Classification

Random Forest

Rin Kawnao (B11015017)

Scikit-learn
Install

```
# 👇 for Anaconda
conda install -c anaconda scikit-learn
```

Scikit-learn
Import

```
1   import numpy as np
2   from sklearn.ensemble import RandomForestClassifier
```

Preparing Data

```
4   np.random.seed(0)
5   X_train_fpath = './HW2/data/X_train'
6   Y_train_fpath = './HW2/data/Y_train'
7   X_test_fpath = './HW2/data/X_test'
8   #output_fpath = './output_{}.csv'
9
10  # Parse csv files to numpy array
11  with open(X_train_fpath) as f:
12      next(f)
13      X_train = np.array([line.strip('\n').split(',')[1:] for line in f], dtype = float)
14  with open(Y_train_fpath) as f:
15      next(f)
16      Y_train = np.array([line.strip('\n').split(',')[1] for line in f], dtype = float)
17  with open(X_test_fpath) as f:
18      next(f)
19      X_test = np.array([line.strip('\n').split(',')[1:] for line in f], dtype = float)
20
```

Check Data

```
22    print(X_train)
23    print(Y_train)
24    print(X_test)
25    print(X_train.shape)
26    print(Y_train.shape)
27    print(X_test.shape)
```

Check Data

X_train

Y_train

X_test

```
[[33.  1.  0. ... 52.  0.  1.]
 [63.  1.  0. ... 52.  0.  1.]
 [71.  0.  0. ...  0.  0.  1.]
 ...
 [16.  0.  0. ...  8.  1.  0.]
 [48.  1.  0. ... 52.  0.  1.]
 [48.  0.  0. ...  0.  0.  1.]]
[1. 0. 0. ... 0. 0. 0.]
[[37.  1.  0. ... 52.  0.  1.]
 [48.  1.  0. ... 52.  0.  1.]
 [68.  0.  0. ...  0.  1.  0.]
 ...
 [38.  1.  0. ... 52.  0.  1.]
 [17.  0.  0. ... 40.  1.  0.]
 [22.  0.  0. ... 25.  1.  0.]]
(54256, 510)
(54256,)
(27622, 510)
```

Learn and Test

```
30    # Select Algorithm
31    clf = RandomForestClassifier()
32
33    # Learn
34    clf.fit(X_train, Y_train)
35
36    # Test
37    Y_test = clf.predict(X_test)
38
39    # DataType float -> int
40    Y_test = Y_test.astype(np.int64)
```

Output
Creating output file

```
48  import csv
49  with open('./HW2/output.csv', mode='w', newline='') as output_file:
50      csv_writer = csv.writer(output_file)
51      header = ['id', 'label']
52      print(header)
53      csv_writer.writerow(header)
54      for i in range(27622):
55          row = [str(i+1), Y_test[i]]
56          csv_writer.writerow(row)
57          print(row)
58
```

- https://ai-kenkyujo.com/programming/how-to-use-scikit-learn/
- https://deepage.net/features/numpy-shape.html
- https://qiita.com/mshinoda88/items/8bfe0b540b35437296bd
- https://bobbyhadz.com/blog/python-no-module-named-sklearn#:~:text=The Python "ModuleNotFoundError: No module,pip install scikit-learn command.
- https://note.nkmk.me/python-numpy-dtype-astype/